The evolution of metanorms: quis custodiet ipsos custodes?

Michael J. Prietula · Daniel Conway

Abstract How are norms maintained? Axelrod (in Am. Political Sci. Rev. 80(4): 1095–1111, 1986) used an evolutionary computational model to proffer a solution: the metanorm (norm to enforce norm enforcement). Although often discussed, this model has neither been sufficiently replicated nor explored. In this paper we replicate and extend that model. Results were generally supportive of the original. Speculations in the original regarding the requirement to link sanctions underlying the metanorm structure were *not* supported, as differentiating punishment likelihoods against defectors from punishment likelihoods against shirkers (non-enforcers of the norm against defection) lead to more efficient and effective sanctioning structures that allowed norm emergence. Replications of the Groups game (two groups differing in numbers and power) generally supported the original reports, but true norms against defection emerged only if sanctioning structures were differentiated, resulting in the Strong group developing a dominant norm against *others* defecting (Metavengeance). That is, when groups are involved with differential power, Metanorms fail unless a more sophisticated sanctioning structure (Metavengeance) is supported.

Keywords Norms · Metanorms · Evolutionary model · Cultural algorithm · Simulation

Norms play essential roles in regulating aspects of social behavior in groups. Norms are viewed as quasi-formal rules and often "instruct strangers and convey to children"

M.J. Prietula (✉)
Goizueta Business School, Emory University, Atlanta, GA 30322-2710, USA
e-mail: prietula@bus.emory.edu

D. Conway
Department of Business Administration, Augustana College, 639 38th Street, Rock Island,
IL 61201, USA
e-mail: DanielConway@augustana.edu

how to behave in specific situations from a particular groups' perspective—norms are models of (situationally) correct behaviors that ostensibly afford some benefit to the referent group (Brown 1995). Norms are found in virtually all human societies (Eibl-Eibesfeldt 1989). When interpreted as guiding standard rules of behavior within an organization, norms influence much of organizational decision making (March 1996). Norms, as opposed to laws in our perspective, are often unwritten and generally do not emerge through an explicit democratic or other formal governing mechanisms; however, the relationship between norms and laws can be intricate and substantial (Ellickson 1991; Horne 2000; McAdams 1997–1998; Posner 2000).

What processes that do (or can) account for norms depends on how one defines norms, the level of specification, and in many cases, the particular function a norm serves in the context of the referent group (Brown 1995) and any particular form is substantially influenced by disparate factors depending on the context (Hechter and Opp 2001).[1] Our interpretation of norms is in line with what are often called *social norms* reflecting their socially-situated (and socially-shared) definition and level of informality, differentiating them from legal and moral norms, conventions, habits and fads (Elster 1989b; Hechter and Opp 2001; Tuomela 1995). In addition to their relative institutional informality, universal acceptance, and (presumed) benefit to the group, another property of social norms is that their violation will bring extra-legal sanctions; that is, violation of a norm is likely to engage some sort of *punishment* from the referent group, directly or indirectly (e.g., Coleman 1990; Dalton 1948; Hackman 1992; Wendel 2001). Thus, the components of acceptable behavior within the group are largely defined and enforced by intra-group processes, and it is presumed that sanctions, in whatever form invoked, have an impact on the decision to deviate from a norm. A norm is not a norm unless there is some form of sanction for its violation (Coleman 1990, Chap. 11; Gibbs 1966; Horne 2001a). By extra-legal direct punishment we eliminate specific legal remedies, such as arrest, prosecutions, and lawsuits; rather, the nature of the sanctions are seen as non-legally enforceable but socially punitive (e.g., shunning, gossip, verbal reprimands) where social capital within the group rises and falls, be it power, reputation, information, or where membership in the group itself is threatened. Indirect punishment can be interpreted as negative feelings (e.g., guilt, shame) that are derived from a social context and serve as sanctions or threats of sanctions, such as an "internalized sense of duty" (McAdams 1997–1998, p. 340). The net effect of any (or all) of these sanctions is assumed to be known to, and calibrated by, all members of the group. In Hardin's (1968) phrasing, this is "mutual coercion, mutually agreed upon."

---

[1]Although many writers conceive of the general nature of social norms, they often disagree on their specific substantive characteristics, how they emerge, the methods of analyzing norms, typologies and so forth, often by disciplines (see Hechter and Opp 2001). For example, Elster (1989a) views social norms as not rational in the sense that they are not followed because of the outcomes they will produce (i.e., they are simply "followed" without regard to any prospect of future reward or loss), but this aspect is disputed by Hardin (1995) who argues that "the push of self-interest might determine very many features of norms, including their forcefulness and their form or structure" (p. 108). It is clear that one can find norms that have no apparent social value (at least to the observer) as well as those that do, and even some that have distinctly maladaptive social values. Examples are often snapshots in the life of a norm and may not indicate its historical role in the group, nor its viability in the future.

It is given that social norms exist, that they often influence behavior and perceptions of behavior, and that they play a broad variety of important social functions within small groups (Marques et al. 2001), within communities and neighborhoods (Ellickson 1991), within professional communities (Wendel 2001), within corporations (Rock and Wachter 2001), within nations (Posner 2000), and even between nations (Tarzi 2002). The focus of this paper is to investigate how a simple structural model of social adaptation and sanctions can account for how cooperation norms emerge, spread, prevail, change, or fail in the culture of a small organization of computational agents. It is a straightforward agent-based game where cooperation matters. The paper is a replication-extension study of the original Axelrod core model, thus addressing an oft-neglected component underlying the valid accumulation of knowledge in the social sciences (Campbell and Stanley 1963). The results conditionally support the original, demonstrating that a two-tiered approach to sanctioning was sufficient to support the emergence of a norm against defection. However, when considering the situation of two groups of differential power (and differing interests), the results are somewhat more complicated.

1 The metanorms game

The approach we take to study norms is based on extending the theoretical and computational work of Axelrod (1986), hereafter referred to as AMG (Axelrod's Metanorm Game), who demonstrated how an evolutionary approach could simulate the emergence, maintenance, and displacement of norms in a group of agents. According to Axelrod, "a norm exists in a given social setting to the extent that individuals usually act in a certain way and are often punished when seen not to be acting in this way" (1986, p. 1097). The evolutionary component was metaphoric as the entities simulated were actually likelihoods of behavioral choices under specific circumstances—behaviors (realized as procedures), not organisms, were evolving. The concept of the social evolutionary process itself (as defined by a genetic algorithm in this case), however, was less metaphorical and more a plausible representation of a societal action and response apparatus underlying the emergence of specific behavioral choices—people tend to copy successful behaviors and tend to avoid unsuccessful behaviors in a group setting when social goals are relevant in a broad variety of contexts (Axelrod 1984; Bandura 1977; Hackman 1992; Kahan 1997). Over time, there is both an exogenous and endogenous "push and pull" of influences and competition for dominance (or even survival) of behaviors in the group that determines the spread and stability of norms.[2]

---

[2]Social pressure to conform is one of the earlier phenomena addressed in modern social psychology (see Festinger et al. 1950) with salient demonstrations well-documented in the literature of classic studies such as Asch (1956) and Milgram (1974). There is also growing evidence that the tendency for copying behaviors in a group setting in general (i.e., imitative learning sans explicit instruction) could be a function of our evolutionary heritage that facilitates the cultural transmission of information (Boehm 1997; Dugatkin 2000; Tomasello 1999), which is related to more general arguments proffering the emergence of culture itself (directly or indirectly) as influenced by components of biological evolution and neuroscience (Boyd and Richerson 1985; Byrne and Whiten 1997;

1.1 Quis custodiet ipsos custodes? (Who watches the watchers?)

Axelrod's solution to a defection problem was demonstrated in the *Metanorms game* where a unique tiered and computational approach structured the influence of norms monitoring defectors (and sanctioning them) from an *N*-person Prisoner's Dilemma game. The Metanorms game first begins with the Norms game. In the *Norms* game, an agent chooses either to defect with gains to itself and losses to the rest of the group of (other) agents, or to not defect (cooperate) thus not harming the group, but also not receiving any gains for itself. If an agent defects, all other agents have a chance to detect that defection and then must make a choice to *sanction* the defecting agent or not, but sanctions incur a cost to the delivering agent. The individualistic payoff structure of the basic game is biased toward defection (egoism), so agents eventually develop a preference to defect at the expense of the group and avoid a preference to sanction those who defect. The defection strategy spreads rapidly within the group and soon becomes stable over the duration of the trials.

There are three characteristics of the Norms game that make it a distinctive approach to the study of norms. First, the agents are in a type of *social dilemma,* which is a decision situation in which "each member of a group has a clear and unambiguous incentive to make a choice that—when made by all members—provides poorer outcomes for all than they would have received if none had made the choice" (Dawes and Messick 2000, p. 111). Individually, each agent in the Norms game prefers to defect, as that would be the dominant strategy if no other agents mattered, based on the individual payoff structure. However, the negative externalities imposed by defection are not a desirable set of outcomes for the group. In fact, if all agents defected each would suffer a loss based on the $N-1$ other agents, so the preferred strategy of defection is a *group deficient* solution.

Second, the substance of the solution to this social dilemma is instituted as an opportunity to impose *sanctions* by members of the same group—a mechanism for norm enforcement. Sanctions are effective solutions to social dilemmas from several perspectives (e.g., Boyd and Richerson 1992; Falk et al. 2005; Voss 2001), but sanctioning generally incurs a cost to the punishing agent as well as the agent that is punished, so the risk of free-riding (not to sanction) can be substantial, therein defining the *sanctioning problem*. As Dawes (1980) points out, "Sometimes, in fact, it is not even possible to avoid a dilemma by reward or coercion, because the costs of rewarding people for cooperating or effectively coercing them to do so exceed the gain the society derives from having everyone cooperate than defect" (p. 175). Indeed, this is exactly what occurs in the Norms game; each agent could invoke a sanctioning strategy, but this strategy does not survive over time. The social mechanism of copying the best performing strategy (as will be explained) dooms the imposition of sanctions, as the agents who do not sanction perform best and those who do sanction quickly learn that sanction does not pay.

Finally, the agents in this simulation have negligible knowledge of social context or contacts; that is, there are no provisions for reputation or contracting, no memories

of prior histories of interactions, no ability for extended reciprocities (beyond the specific episodic opportunities to be described), no ability to look forward (no shadow of the future), no ability for backward induction, and no abilities to distinguish one agent from another. The agents in this simulation are quite simple as are their mechanisms for selecting strategies and behaviors. They are minimally cognitively and socially rational (Carley and Newell 1994) seeking only to earn benefits for themselves and to fit in the group (i.e., be acculturated) by copying the most successful behaviors—by being conformists (Henrich and Gil-White 2001).

So how can a norm against defection arise within this group of egoists? The problem resides not in the use of sanctions, but in the *choice* to (not) impose sanctions. In effect, this is a second-order free rider (or public goods) problem (Coleman 1990, pp. 270; Oliver 1980). The answer proposed by Axelrod is based on defining a *second* type of norm—a *metanorm*—that specifies that those agents who do not sanction defectors should be sanctioned themselves. The concept of a metanorm as a solution to the second-order free rider problem has been proposed in a variety of forms (Axelrod 1986; Coleman 1990); however, the metanorm solution has also been criticized as it "pushes the problem to a higher order" (Henrich and Boyd 2001), it involves substantial circularity in reasoning regarding commitment and monitoring where "the process unravels at both ends" (Ostrom 1990), does not seem to be often documented (Elster 2006), or it is restricted to small, close-knit communities (Lichbach 1996). Nevertheless, work by Boyd and Richerson (1992) suggests that the use of sanctions against defection under certain circumstances can yield what they call *moralistic strategies* "which cooperate, punish noncooperators, and punish those who do not punish noncooperation can be evolutionary stable" (p. 173). Consider Coleman's (1990) story of how such metanorms can arise in the London financial district to mitigate against investment bankers to defect from a code of ethics, which impacts the entire community:

> The first norm must be something like this: "Do not engage in transactions with a party who has violated the code of ethics." And that norm must be backed up by sanctions, which in such an informal community may necessitate another norm, something like the first: "Do not engage in transactions with a party who engages in transactions with a party who has violated the code of ethics." … Such a normative system is difficult to maintain unless the community is very close and very homogeneous in interests. (p. 116)

This is the substance of the Metanorms game. Note that in Kollock's (1998) fundamental framework of solutions to social dilemmas, it is a combination of *strategic* (social learning) and *structural* (sanctions) remedies. The Metanorms game extends the Norms game by a second round allowing agents to evolve a strategy to detect and punish an additional type of behavior: agents that do not punish defectors. In the Metanorms game (and contrary to the Norms game) a *norm* against defection does emerge: individualism is reduced and cooperation is increased. Thus, only an additional *single* level of monitoring-the-monitors is sufficient to generate the spread, and sustain the existence, of a behavioral norm to quash levels of individualistic behaviors (defections) that could hurt the group. Contrary to predictions of infinite regress and *N*th order free-rider issues, cooperation emerges from a solution of using two

norms that are based on the *same* pay-off structure. It was not the pay-off structure that was allowed to vary, but the level of tolerance for defections in general (for defectors and for shirkers) *across* group members as a replicated strategy *based* on the pay-off structure. In a sense, it is shows that norms can emerge if the group values defection from enforcement the same way it values defections from cooperation, and this valuation is the same within the group. Thus, a simple sanction structure provided a solution was sufficient to account to the emergence of the AMG norm.

2 Replication and extensions

This paper explores three elements of metanorms within the context of the original model. First, we *replicate* the primary study of the AMG of the Norms and Metanorms games, asserting the following hypothesis:[3]

Hypothesis 1 The algorithm description in AMG sufficiently describes the processes that account for the emergence of a stable norm against defection.

Second, we test the *correlated vengefulness hypothesis* (our title) that was proposed in the original AMG study, which asserts that for metanorms to be effective, there needs to be an association between the sanctions toward a socially undesirable behavior (defection) and the sanctions toward those not punishing that behavior (shirkers). As Axelrod posited:

> The trick, of course, is to link the two kinds of vengefulness. Without this link, the system could unravel. An individual might reduce the metavengeance level while still being vengeful and then later stop being vengeful when other stopped being metavengeful. (1986, p. 1102).

In the original Metanorms game, this was realized by having the two sanctioning decisions based on the value of a *single* strategic construct (vengefulness toward all). What this meant was that any given agent could not selectively differentiate sanctioning probabilities between the two contexts: agents that refuse to sanction defectors and agents that refuse to sanction those agents. On the other hand, allowing the two choice utilities to vary independently may lead to a more efficient sanctioning structure. We test this hypothesis by *decoupling* the two types of sanctioning contexts so that each can strategically evolve separately in the Metanorms game.

Hypothesis 2 The norm against defection is based on a link between sanctions against a defector and sanctions against a shirker (of norm enforcement).

---

[3]The original model used 5 replications and 100 generations per run. Our tests demonstrated that these are insufficient for stable results. We increased replications to 100 per condition to increase the basic power (Cohen 1988) and increased the length of the simulation/generations allows us to address findings that question the duration strategy viability in dynamic games where any particular strategy has a non-zero chance of extinction (e.g., Young and Foster 1991).

Finally, we re-examine the *Groups game* (our title) played in AMG. This was a version of the Metanorms game but composed of two different social groups of agents that differed in group size and power (i.e., ability to impose sanctions)—the Strong group versus the Weak group. In the original study, group affiliation mattered for both the Norms and Metanorms games. Defection in the Norms game only hurt members of the opposite group, and defectors could be sanctioned only by members of the opposite group, reflecting long standing research on in-group bias under a remarkably broad variety of conditions for choice (e.g., Allport 1954; Sherif et al. 1954/1988). On the other hand, violations of in-group norms also results in sanctions (Eidelman and Biernat 2003; Shinada et al. 2004) and the impact of such sanctions, directly and indirectly (as a message) can be differential and substantial (Bernhard et al. 2006; Williams 2007). In the sanctioning structure of the Metanorm games, in-group sanctioning occurred when shirking was detected—in-group members did not punish members of the group who refused to punish defectors of the other group. Though little research has explicitly explored the link between such group-metanorm linkages, there exists evidence that such behavior exists, in part, to help solidify groups (e.g., Horne 2001b), and more indirect results as Hornsey et al. (2002) suggesting a sensitivity effect with their group research: criticism from in-group members were tolerated much better than criticism from out-group members. The findings of AMG suggested that metanorms were required to alleviate the shirking that emerged even in the Strong group and thus allow the emergence of a stable norm against defection. The original findings reported were as follows:

> Resistance to punishment and increased size can help a group, but only if there are metanorms. Without metanorms, even members of the stronger group tend to be free riders, with no private incentive to bear enforcement costs. This in turn leads to low vengefulness and high boldness in both groups. When metanorms are added, it becomes relatively easier for the strong group to keep the weak group from being bold, while it is not so easy for the weak group to keep the strong one from defecting. (Axelrod 1986, p. 1003)

Thus, without metanorms there should be high defections in both groups. The inclusion of metanorms should reduce the defections in the weak group but less reduction should occur in the strong group. We replicate the original study (additionally increasing the replications and lengthening the generations to discern sensitivities) offering the following hypothesis:

Hypothesis 3 The norm against defection between groups of differential power emerges when groups engage in metanorm sanctioning, where the Strong group dominates the Weak group.

We then extend the model in the following two ways: (1) *decoupling* the two types of sanctioning contexts as described in Hypothesis 3, and (2) adjusting *group affiliation* in sanctioning decisions. The former manipulation explores whether decoupling the sanctioning structures can more efficiently support norm emergence or not. This is the testing of Hypothesis 2 examining the impact of a dual utility structure for sanctioning under Metanorms, thus this hypothesis is presumes the emergence of a norm

against defection only under the Metanorm sanctioning structure, with the impact of influence following the pattern of behavior revealed by the prior hypothesis:

Hypothesis 4 The norm against defection is based on a link between sanctions against a defector and sanctions against a shirker (of norm enforcement), and will vary according the pattern found in Hypothesis 2.

3 Reproducing the basic results

The first step was to reproduce the primary study of AMG. This involved running the Norms game and the Metanorms game with the parameters used in the original paper (see Fig. 1). The algorithm is presented in the Appendix and the primary conditions and parameters are shown in Table 1. In AMG, twenty agents were faced with the following three decision situations: (1) an *N-person Prisoner's Dilemma* decision as the core, where each agent, upon its turn, chooses whether to defect or not, resulting in differential gains to the agents, (2) a basic *Norms game* decision, where each agent chooses whether or not to punish a defector, if that defector is observed, and (3) a *Metanorms game* decision, where each agent chooses whether or not to punish a shirking agent (an agent who did not punish an observed defector). Agents in AMG behave in each of these decision situations according to their particular decision strategy. Strategies are defined in a 2-dimensional construct space where one axis is *boldness*—likelihood to defect in the *N*-person Prisoner's Dilemma decision. The other axis is *vengefulness*—the likelihood of punishing an agent detected for either defection or non-punishment. For every agent, boldness or vengefulness can take on one of eight values, ranging from 0/7 to 7/7.

As noted, the Norms and Metanorms games employ simplifying assumptions that distinguish it from similar games. The games are played among agents who are equally likely to encounter opportunities to defect and to observe defections, as

Table 1 Initial conditions and parameters for computational experiments: AMG, Groups games

Parameter	Value
Number of agents	AMG: 20, Groups: 30
Initial boldness, vengefulness	Random
Defection opportunities per round	4
Number of generations	100 for replication, 1000 for extension[a]
Number of replications	5 for replication, 100 for extension[a]
Probability of defector detected, s	Exogenous random (uniform) [0, 1]
Probability of shirker detected, s'	Exogenous random (uniform) [0, 1]
Benefit from defection, T	3
Cost to group, H	-1
Norm enforcement cost, E	-2
Punishment cost to defector, P	-9

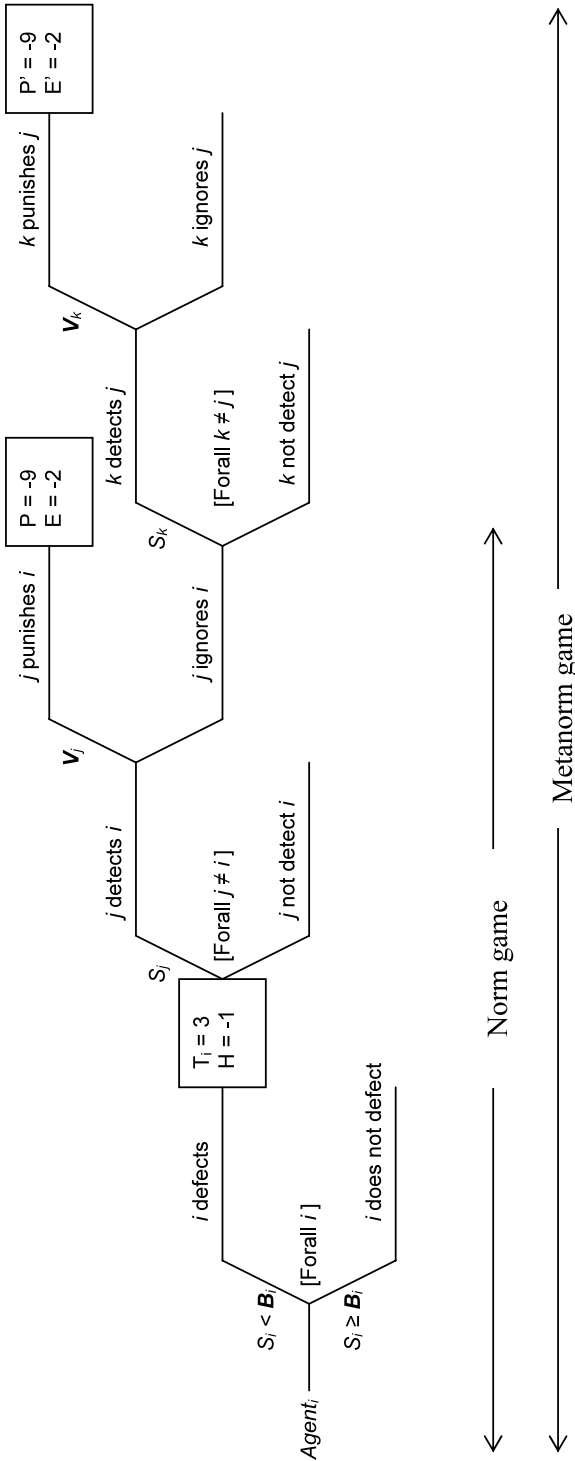[a]Convergence occurs sufficiently with these extension values (variation within 5%)

Fig. 1 Structure of basic Norms and Metanorms games

both detection and defection rely on the same exogenous probability value randomly drawn for that particular event. What differs is each agent's decision of what to do when such opportunities are presented based on the values of their strategy constructs (boldness, vengefulness). The dynamics for change in an agent's strategies are based on a simple replication algorithm of the most successful strategies in the population. The Metanorms game was played with the same constraints and parameters as the Norms game (i.e., constant population at twenty agents, 100 generations, five replications). Results indicated that stable solutions occurred within 1000 generations, and sufficient power was obtained with 100 replications each, so these underlie the reported findings (details available from the authors).

3.1 Results

Analysis of the values on the 1000th generation across replications supported H_1: the incorporation of Metanorms significantly reduced the average Boldness values ($F(1, 198) = 88191.46$; $p < 0.000$) leading to significantly less defections ($F(1, 198) = 56923.52$; $p < 0.000$). Metanorms indeed inhibit defections. The dynamics of how mean population values change (boldness and vengefulness scores) are illustrated by a single run from each game shown in Fig. 2. The plots of the final values are shown in Figs. 3 and 4 across 1000 replications each for the Norm and Metanorms games respectively.
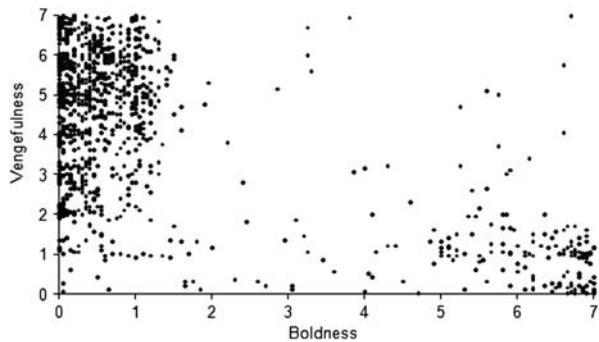

Fig. 2 Example traces of average population strategy values over 100 generations
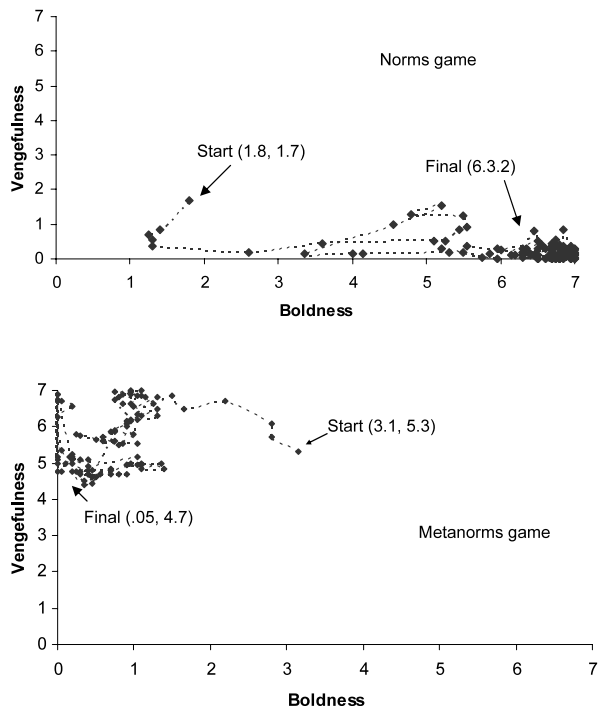
Fig. 3 Final strategy values for
Norms game plotted over 1000
replications
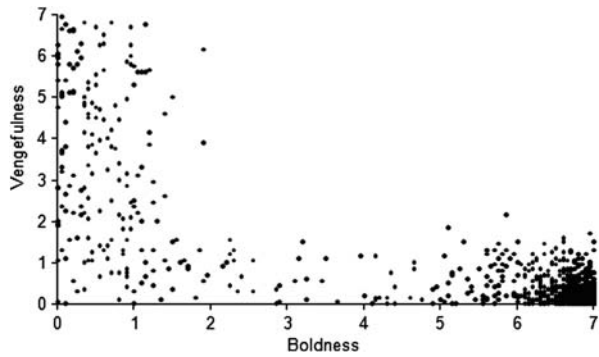
Fig. 4 Final strategy values for
Metanorms game plotted over
1000 replications

4 Correlated vengefulness (decoupling metavengeance)

A set of simulations was run that allowed a third strategy dimension—Metavenge-
ance—to evolve independently. All other constraints and procedures were carried out
as the basic Metanorms game, with 100 replications and 1000 generations.

4.1 Results

The results indicate that by decoupling the sanctioning strategies does not prevent
the emergence of the norm against defection, thus not supporting Hypothesis 2.
There were no significant differences between the two Metanorm conditions on av-
erage Boldness ($F(1, 198) = 1.21$ ns), average Vengefulness ($F(1, 198) = .41$ ns),
or average Defection levels ($F(1, 198) = .83$ ns). However, in the decoupled condi-
tion the levels of Metavengeance were significantly *lower* than the levels of evolved
Vengeance (Wilcoxon, $z = 7.82$, $p < .001$) indicating less sanctioning efforts were
required to enforce the Metanorm than the Norm.

5 The groups game

The Groups Game consisted of a slight variation to the Metanorms game where two
different agent groups were defined and differed in number and power reflecting ad-
vantages of one group over the other. The first group (Strong) consisted of 20 agents

and the second group (Weak) consisted of 10 agents, where both group population size's were held constant. Differences in power were reflected in the patterns of punishment and the relative impact of punishment on scores.

For the Norms game, a defection by a member of one group only hurt the members of the *other* group, so detected defectors are punished only by the members of the *other* group. For the Metanorms game, an agent shirking the punishment of a defector (from the other group) is only punished by members of the *same* group to which it belongs. Additionally, Strong agents punishing Weak defectors retained the original punishment score of $P = -9$, but Weak agents punishing Strong defectors was lowered to $P = -3$. However, for each generation, Strong agents learned only from Strong agents and Weak agents learned only from Weak agents, where strategies were evaluated and spread solely within groups. All other values and procedures remained the same, and vengefulness values were *coupled* (as one Vengefulness component) as in the original study. The analysis was based on the 1000th generation values over 100 replications. The Groups game was played under three conditions: the Norms game, the Metanorms game, and the Metanorms game with a decoupled Metanorm structure (metavengeance).

5.1 Results

In the Norms game, there was no impact of groups and no norm emergence (leftmost column, Fig. 5). There was an overall main effect of Metanorms in the Groups Game where Boldness decreased significantly ($F(1, 398) = 151.76$; $p < .001$) and
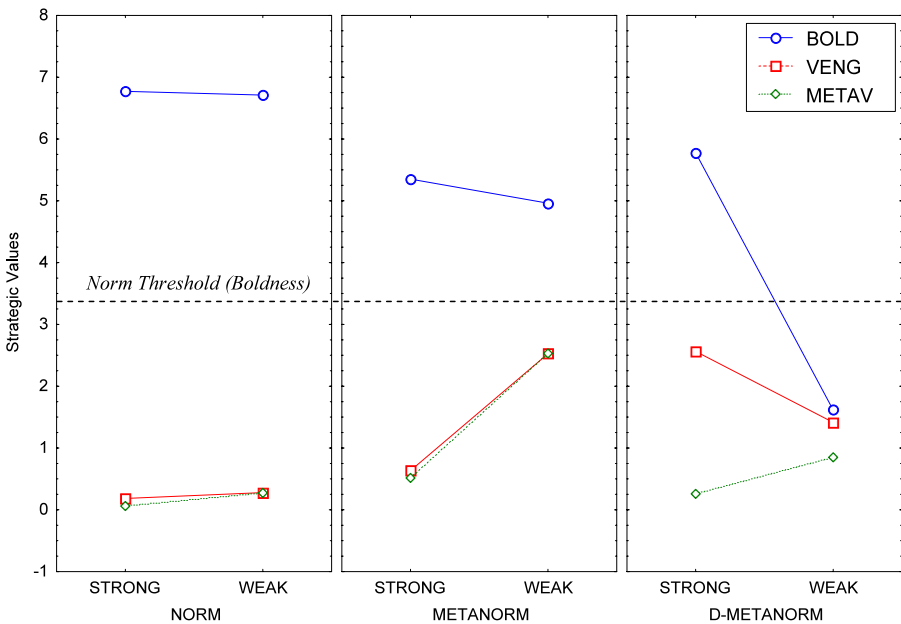

Fig. 5 Boldness (BOLD), Vengefulness (VENG) and Metavengefulness (METAV) scores for each group (Strong, Weak) by game (Norm, Metanorm, Decoupled Metanorm)

Vengefulness increased significantly ($F(1, 398) = 140.39$; $p < .001$) as shown in the middle panel of Fig. 5. However, the *overall* Boldness levels under the Metanorms game did not reach levels sufficient to be described as effectively realizing the norm against defection (defined as less than the mid-value of the scores, 3.5). So who were defecting? An analysis of the defections by Group revealed that the Strong group defected at significantly higher proportions than the Weak group under both the Norms and Metanorms games ($F(1, 398) = 761.08$; $p < 0.001$), supporting Hypothesis 3. But this must be interpreted in terms of dominance as was done in the original discussion—Metanorms effectively instituted a norm against *relative* defection for the Weak group, while the Strong group could defect with impunity, but a "true" norm against defection was not emergent in either group—rather, a dynamic struggle between groups of different power emerged with little overall control against defection. Consider Fig. 6 (upper) of the Metanorms game. Both weak and strong groups have high levels of Boldness, with the Strong group succeeding with even lower levels of within-group punishment than the weak group. When groups are involved, Metanorms fail.

However, by decoupling the Metavengeance from Vengeance, things changed substantially (right panel, Fig. 5). Relative to Metanorms levels, there was a significant interaction between the use of Metavengeance and defections ($F(1, 396) = 53.32$; $p < 0.001$), with a post-hoc analysis showing that defections by the Strong group remained essentially the same (Tukey HSD, ns) while defections by the Weak group decreased significantly (Tukey HSD, $p < .001$).

Fig. 6 Sample runs of Groups game for Metanorm game (*upper*) and Metanorm game with Metavengeance (*lower*) of 1000 generations
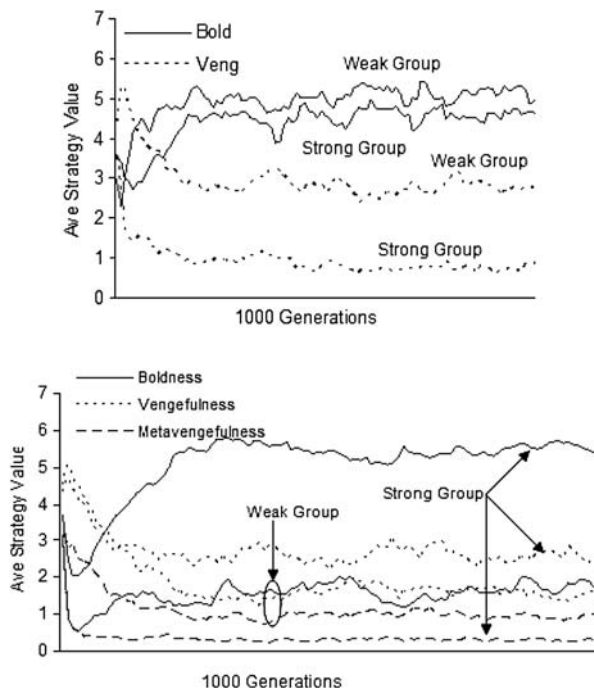
Recall that we define the norm against defection when Boldness levels are less than the midpoint of the possible strategic values (3.5), and this is indicated as the "norm threshold" line in the graph. As noted, under the Norm and Metanorm sanctioning structures, Boldness levels do not fall below the norm threshold. The rightmost column graph shows the impact of separate Metavengeance structure. First, there is an interaction between sanctioning structure and Boldness ($F(1, 396) = 102.43$; $p < .001$) where Boldness values for the Weak group drops significantly from the Metanorm condition (Tukey HSD, $p < .001$) and this drop takes it below the norm threshold of 3.5 (mean $= 1.61$). However, the Boldness level for the Strong group remains unchanged (Tukey HSD, ns). Insight to this can be gained by looking at the Metavengeance structure in Fig. 5 (last column) and Fig. 6 (lower). The Strong group has significantly increased Vengeance levels (from the Metanorm condition) against the Weak group (Tukey HSD, $p < .001$) and has a substantially lower Metavengeance level to sustain the in-group metanorm (Tukey HSD, $p < .001$). As the defection rates for the Strong group dominated, there was an emergence of a "norm against *others* defecting." Consequently, the Strong group evolved high Boldness scores (and could defect with impunity), moderate Vengefulness scores (to control defection of the Weak group), and low Metavengefulness scores (little value gained for punishing their own shirkers, except when necessary). On the other hand, the Weak group was essentially forced into docility with lower levels of strategy values on all three dimensions. Once the sanctioning structures were decoupled, the Stronger group could adapt more effectively and efficiently (via the sanctioning structure mechanism) to exploit its advantage over the Weaker group in order to quash defectors. This means that a Metavengeance sanction structure allows a true norm against (others) defection to emerged and be sustained by the Strong group against the Weak group and Hypothesis 4 is not supported. When groups are involved, Metanorms can function only if a more sophisticated sanctioning structure (i.e., Metavengeance) is supported.

6 Discussion

The results of the present study support, clarify, and extend the findings of the theoretical model and simulation of Axelrod (1986) addressing the dynamics of norm emergence from the perspective of agent-based modeling of social groups (Axelrod 2006; Davis et al. 2007; Harrison et al. 2007). The basic outcomes of the AMG simulation were replicated. Without metanorms, a norm against defection cannot survive; with a single metanorm structure, a norm against defection can evolve and survive in a group where parameters suggest "dependence" (see also Horne 2004). The examination of the *correlated vengefulness hypothesis* by decoupling Vengeance from Metavengeance clarified the suspected correlation between the two kinds of vengefulness. Axelrod proposed that a link between the two was required (and realized by a single vengefulness construct) in order to avoid additional free-rider problems. In fact, this link is *not* required. Two distinct norms emerge and are maintained by two distinct sanctioning structures. The Vengeance level of the decoupled norm is equivalent to that of the coupled Norm, but the Vengeance level of the decoupled metanorm is approximately 40% lower. This affords a substantially more efficient

adaptation for enforcing norms. Furthermore, the levels of vengefulness for norms were substantially higher than those for metanorms, which seems counter to Axelrod's (1986) speculations: "The types of defection we are most angry about are likely to be the ones whose toleration also makes us angry" (p. 1103). It is possible, but it may not be necessary.

This has been generally seen as a counter-intuitive result, but this model result reflects a *structural* explanation of norm emergence. The key for why metanorms work (in both coupled and decoupled situations) is found in the nature of how strategies are spread *within* the group. In essence, the agents are all equivalently *docile*—the best performing strategies are readily adopted (replicated) by any given agent, as no agent is predisposed (or committed) at any time to any particular strategy. Furthermore, defection strategies (Boldness) and sanctioning strategies (Vengefulness, Metavengefulness) are *orthogonal*, so sanctioning costs can be absorbed even by "cheaters" (Eldakar et al. 2007; Nakamaru and Iwasa 2006) and avoiding the need to invoke (or explain) what Coleman (1990) calls a *heroic sanction* which is "a sanction whose total effect occurs through a single agent" (p. 278). The pay-off structures (of the individual agents) and the dynamics of the group interactions achieve a *cultural equilibrium* of values defining the metanorm, which appears to also occur under different metanorm evolutionary models (Kendal et al. 2006) and is plausible within a general voting model of society (Hurwicz 2008).

Finally, the replication of the Groups game presented interesting results. Recall that norms in these games were defined according to groups, where a norm against defection actually meant a norm against the *other* group defecting. Whenever alliances among agents occurred in the Norms game, a norm against defection could not emerge in either group, with accordingly high defection rates. However, if alliances were *prevented* in the Norms game, *both* groups could develop a norm against defection and overall defection rates were quite low and approached levels found in the condition without groups. Thus, the overall defection rate (across both groups) was minimized and a "population beneficial" solution was reached. The problem, then, centered on situations where alliances are made in the Norms game. The solution to this was the decoupling of the vengeance scores, but the results were not population beneficial. When both groups can evolve separate norms and metanorms, the Strong group rapidly adopts a stable "norm against defection by the weak" and the Weak group becomes docile and ceases to defect at significant levels. The Stronger group also developed higher vengeance against the Weak group defections that it did to its own group defections (shirking). The dominance of the Weak group by the Strong one is complete.

The studies presented in this paper were based on small extensions to the AMG model. We envision five areas of further exploration related to this work that may provide additional insight into the limits and value of this type of metanorm structure.

First, manipulations can be made in the *pay-off structures* of the agents and the resultant preference ordering among alternatives. Systematically varying the values of the pay-off structures but retaining the inequalities can provide insight into the sensitivities of the results of the parameter space within the general form of the games; varying certain inequalities can determine the impact of the metanorms structure on varies game forms. On the other hand, varying the within-game values implies that

games can change dynamically. For example, the current structure assumes that defection and shirking are punished equally (in Fig. 1, $P = P' = -9$) and incur an equal enforcement cost ($E = E' = -2$). Research indicates that as norm enforcement costs rise, sanctioning of defections decrease, but metanorms can become stronger (Horne and Cutlip 2002) especially under proc-social influence (Horne 2007). Apparent "costly" punishment is, in part, absorbed by group identification or even more basic mechanisms (de Quervain et al. 2004; Knutson 2004). Furthermore, it may be necessary to expand to more complex incentive structures to account for sanctioning failures in cooperation (Houser et al. 2008).

Second, there are elements of the *cultural algorithm* that can be explored. For example, the current algorithm assumes that all agents try to "imitate the most successful" performing strategies, but "imitate the most common" is also a plausible strategy as both can be interpreted as heuristics for social learning (Boyd and Richerson 1985). Once adopted, the current model assumes that norms can change rapidly and effortlessly, like fads and informational cascades (see Bikhchandani et al. 1998, 1992). Indeed, norms can change rapidly (Axelrod 1986), but on the other hand norms can also be quite resistant to change beyond the structural sanctioning support afforded by the model, and once a norm becomes entrenched, the switching costs to society can result in a suboptimal norm "trap" (cf., Posner and Rasmusen 1999). If a norm is internalized (Axelrod 1986), it generally reflects something of a shift from external sanctions to a self-sanctioning structure, thus perhaps reducing the need for (or level of) sanctioning for norms or metanorms. This could be easily modeled by defining a *docility index* for any given agent that reflects an agent's susceptibility to strategy value adoption (or change) based on, for example, an inverse function of the time a norm is held—the longer a norm is held, the more resistant it is to change. Resistance itself may be considered a culturally plausible construct to model.

Third, such cultural algorithms (or any group-based sanctioning or conformity phenomenon) as defined here can be susceptible to *group size effects* (e.g., Agrawal and Goyal 2001; Borrett and Patten 2003; Friedrichs and Blasius 2003; Stang 1976), though the effects are often neither simple nor straight-forward (e.g., Barnir 1998; Carpenter 2006; Isaac and Walker 1988; Marwell and Ames 1979; Rapoport 1988; Tata and Anthony 1996). As we have noted, some of the criticisms against this type of metanorm structure have argued that it is plausible only within "small homogenous groups". Of course, there is not absolute definition of "small" so it necessary to see if there are group size effects and how they are manifested under varying conditions and assumptions. For example, the basic reciprocity explanation for altruism (under standard evolutionary theory) is constrained by the number of individuals who are likely to interact (Boyd and Richerson 1988). In addition, increasing the size would impact the detection level (s in Fig. 1 for monitoring agents) reflecting a common finding that vigilance varies inversely with group size (Dunbar et al. 2002) as well as the perception of risk in specific cultures (Ho and Leung 1998) and other interesting hypotheses concerning human networking size in general (e.g., Hill and Dunbar 2003).

Fourth, the cultural algorithm used is based on similar agents; therefore consideration of groups of *heterogeneous agents* is an interesting area of expansion. What we mean by agent heterogeneity is a fundamental different property or set of properties of

agents and not simply the difference in particular property values an agent can assume as the game. For example, in the prior discussion one could define that agents differ in their docility index not also as a function the experiences of the group (i.e., number of generations a strategy value set is held), but also in a discount/acceleration factor that one agent is fundamentally more (or less) docile than another. Similarly, in the cultural algorithm some agents can be more influential than others (De Cremer 2002). Epstein (2001) proposed an evolutionary model of norms that combined aspects of a cultural algorithm based on proximity with an individual agent property of cognitive effort. Bowles and Gintis (2004) as well as Kurzban and Houser (2005) demonstrate how multiple types of reciprocators can emerge and exist in a group that sustains co-operative norms. Interestingly, Cinyabuguma et al. (2006) demonstrate how second-order punishment (i.e., metanorms) can mitigate the impact of a commonly observed perverse behavior of a subset of agents (i.e., punishing high contributors) in voluntary contributions games.

Finally, this study provided insight into how the Groups game and cultural algorithm properties interacted to generate norms, metanorms, and dominance of one group over another. This provides an initial thrust into exploring how such a metanorm architecture can fit into (and perhaps help explain) how group norms emerge, prevail, change, or fail in the *cultures of competing agents* and even how *different cultural groups are formed*. Anthropology has argued that most of the variation between groups is based on cultural differences (Henrich and Boyd 1998), including costly punishment as defined in this paper (Henrich et al. 2006), and those cultural differences (realized as behaviors at the group level) can be considered units of adaptation in multilevel selection theory (Wilson and Sober 1994). Based on this theory, analysis (or simulation) of a single group incorporating the *cultural architecture* defined in this paper (i.e., adaptive components of norms, set of agents with payoff-structures, cultural adoption algorithm, sanctioning mechanism), can yield norm stability in virtually any behavior (Boyd and Richerson 1992). Nevertheless, only between-group selection (i.e., competitiveness in adaptation between groups) can favor social norms that realize functionally adapted groups (Wilson and Sober 1994) and the sanctioning structures have distinct implications for competitive advantages at the group level (e.g., Gürerk et al. 2006). Therefore, it would be interesting to incorporate the metanorm architecture in multiple competing groups, but include alternate sets of norms to selectively enforce or ignore. This could be expanded by including what Heckathorn (1990) called *collective sanctions*, where not only a particular individual would be sanctioned (e.g., for defecting), but the entire group to which the sanctioned individual was a part would also be sanctioned. May the best group win.

Although much can be done to elaborate the original Axelrod model, we should keep in mind a quote by Axelrod (1987) on this specific issue:

> . . . simplicity of theory is always preferable to needless complexity. Nonetheless, society is a stubbornly complex system, and a good model of its dynamics will necessarily be, in some respects, complicated. The norms game, as extended over the generations, is a model of cultural evolution that can readily accommodate this need, a model that is itself open to endless evolution (p. 51).

May the best theory win.

Appendix

Norms game The basic Norms game was run as described in AMG and is described as follows. The *replication algorithm* involves reproduction, crossover, and mutation (Goldberg 1989). Reproduction involves categorizing agents with high payoff scores (at least one standard deviation above the population mean), low payoff scores (less than one standard deviation below the population mean), and mid-scoring agents (between the two extremes) in order to determine what behaviors will be spread (reproduced) in the group. Crossover involves randomly selecting two high scoring agents and randomly duplicating their Boldness and Vengefulness scores (e.g., Boldness of one agent and the Vengefulness of the other), thus spreading their (altered) strategy values to two agents. Mid-scoring agents also cross-over their strategies, but they are passed on to only one agent in a new round. Strategies of low-scoring agents are not replicated. A mutation rate (1%) is applied allowing for random changes in the population strategies as a final step. Note that the incorporation of mutation not only infuses a random component to the algorithm, but also guarantees that no strategy will meet extinction.

1. A population is defined as twenty agents and that population is kept constant. Initial strategy values (boldness, vengefulness) for each agent are selected randomly from integers over the interval [0, 7]. Each agent has a *score card* that tallies its performance value resulting from subsequent payoff events described below.
2. Each of the twenty agents is randomly selected in turn. An agent A_i, upon its turn, is presented with four consecutive situations, or cases, that offer a defection opportunity. For each case, an agent decides to defect if its chance of being seen, s, is less than its current boldness level (B_i), where s is an exogenous parameter drawn from a uniform distribution between 0 and 1.
   2.1. If agent A_i decides *not* to defect, there is no payoff or punishment involved and the case is concluded.
   2.2. If agent A_i decides *to* defect, the defector receives a payoff ($T = 3$) and all other agents receive a punishment ($H = -1$). In addition, the defecting agent A_i bears the chance of being seen, also set at s, by one or more of the other agents in the group.
   2.3. If another agent A_j ($j \neq i$) detects the defection of A_i, then A_j decides to punish A_i with a likelihood based on the current level of A_j's vengefulness (V_j).
      2.3.1. A choice to punish (i.e., enforce a norm against defection) results in an *enforcement cost* to the punishing agent ($E = -2$) as well as a *punishment cost* to the offending agent ($P = -9$).
      2.3.2. A choice not to punish (*shirking*) incurs no costs to either agent.
   2.4. Play the *Metanorm game* (this step is skipped when playing only the Norm game).
3. How the successful strategies are then spread throughout the population agents is then determined by a genetic algorithm, whereby the values of the constructs of the most successful strategies a higher likelihood of be transmitted to (or adopted by) other agents in the population, than do the less successful strategies.

4. Steps 2 and 3 are repeated for 100 generations, with the final strategy values for each agent being the final (i.e., terminally evolved) strategy.
5. Steps 1 through 4 are repeated five times, defining the five replications of the original AMG simulation.

Metanorms game   The Metanorm game is played when a defection occurs (see the Norms game section) and is described under the algorithm as Step 2.4:

2.4. *Metanorms game*. If in Step 2.3 an agent A_j *detects* the defection of A_i, and A_j decides to punish A_i, the Metanorm game is not applied. However, if A_j decides *not* to punish A_i, then
   2.4.1. An agent A_k ($k \neq j$ or $k \neq i$) detects with probability s the lack of norm enforcement by agent A_j. A_k then decides to punish (i.e., enforce a metanorm) based on the current vengefulness level of A_k (V_k).
   2.4.2. A decision to punish results in an *enforcement cost* to the punishing agent A_k ($E' = -2$) as well as a *punishment cost* to the offending agent A_j ($P' = -9$).
   2.4.3. A choice not to punish incurs no cost to the agents.

References

Agrawal A, Goyal S (2001) Group size and collective action: Third-party monitoring in common-pool resources. Comp Political Stud 34(1):63–93
Aiello LC, Wheeler P (1995) The expensive tissue hypothesis. Curr Anthropol 36:184–193
Allport G (1954) The nature of prejudice. Addison-Wesley, Cambridge
Asch S (1956) Studies of independence and conformity: I, A minority of one against a unanimous majority. Psychol Monogr 70(9):1–70
Axelrod R (1984) The evolution of cooperation. Basic Books, New York
Axelrod R (1986) An evolutionary approach to norms. Am Political Sci Rev 80(4):1095–1111
Axelrod R (1987) Laws of life: How standards of behavior evolve. Sciences 27:44–51
Axelrod R (2006) Agent-based modeling as a bridge between disciplines. In: Judd K, Tesfatsion L (eds) Agent-based computational economics. Handbook of computational economics, vol 2. North-Holland, Amsterdam, pp 949–1011
Bandura A (1977) Social learning theory. Prentice-Hall, Englewood Cliffs
Barnir A (1998) Can group- and issue-related factors predict choice shift? Small Group Res 29(3):308–339
Bernhard H, Fischbacher U, Fehr E (2006) Parochial altruism in humans. Nature 442:912–915
Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. J Political Econ 100(5):992–1026
Bikhchandani S, Hirshleifer D, Welch I (1998) Learning from the behavior of others: Conformity, fads, and informational cascades. J Econ Perspectives 12(3):151–170
Boehm C (1997) Impact of the human egalitarian syndrome on Darwinian selection mechanics. Am Nat 150(Suppl):S100–S121
Borrett S, Patten B (2003) Structure of pathways in ecological networks: Relationships between length and number. Ecol Model 170(2/3):173–185
Bowles S, Gintis H (2004) The evolution of strong reciprocity: Cooperation in heterogeneous populations. Theor Popul Biol 65:17–28
Boyd R, Richerson P (1985) Culture and the evolutionary process. University of Chicago Press, Chicago
Boyd R, Richerson P (1988) The evolution of reciprocity in sizable groups. J Theor Biol 132:337–356
Boyd R, Richerson P (1992) Punishment allows the evolution of cooperation, (or anything else) in sizable groups. Ethol Sociobiol 13:171–195
Brown D (1995) When strangers cooperate: Using social conventions to govern ourselves. Free Press, New York

Byrne R, Whiten A (1997) Machiavellian intelligence. In: Whiten A, Byrne R (eds) Machiavellian intelligence II: Extensions and evaluations. Cambridge University Press, Cambridge, pp 1–23

Cacioppo J, Visser P, Pickett C (eds) (2006) Social neuroscience. MIT Press, Cambridge

Campbell D, Stanley J (1963) Experimental and quasi-experimental designs for research. Rand McNally, Chicago

Carley K, Newell A (1994) The nature of the social agent. J Math Sociol 19(4):221–262

Carpenter J (2006) Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. Games Econ Behav 60:31–51

Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Academic Press, Hillsdale

Coleman J (1990) Foundations of social theory. Harvard University Press, Cambridge

Cinyabuguma M, Page T, Putterman L (2006) Can second-order punishment deter perverse punishment? Exp Econ 9:265–279

Dalton M (1948) The industrial ratebuster: A characterization. Appl Anthropol 7:5–18

Davis J, Eisenhardt K, Bingham C (2007) Developing theory through simulation methods. Acad Manag Rev 32(2):480–499

Dawes R (1980) Social dilemmas. Ann Rev Psychol 31:169–193

Dawes R, Messick D (2000) Social dilemmas. Int J Psychol 35(2):111–116

De Cremer D (2002) Charismatic leadership and cooperation in social dilemmas: A matter of transforming motives? J Appl Soc Psychol 32(5):995–1016

de Quervain D, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E (2004) The neural basis of altruistic punishment. Science 305:1254–1258

Dugatkin L (2000) The imitation factor. Free Press, New York

Dunbar R, Cornah L, Daly F, Bowyer K (2002) Vigilance in human groups: A test of alternative hypotheses. Behavior 139:695–711

Eibl-Eibesfeldt I (1989) Human ethology. Aldine de Gruyter/University Press, New York

Eidelman S, Biernat M (2003) Derogating black sheep: Individual or Group protection? J Exp Soc Psychol 39:602–609

Eldakar O, Farrell D, Wilson D (2007) Self punishment: Altruism can be maintained by competition among cheaters. J Theor Biol 249:198–205

Ellickson R (1991) Order without law: How neighbors settle disputes. Harvard University Press, Cambridge

Elster J (1989a) The cement of society: A study of social order. Cambridge University Press, Cambridge

Elster J (1989b) Social norms and economic theory. J Econ Perspectives 3(4):99–117

Elster J (2006) Altruistic behavior and altruistic motivations. In: Kolm S-G, Ythier J (eds) Handbook of the economics of giving, altruism, and reciprocity, vol 1. Elsevier, Amsterdam, pp 183–223

Epstein J (2001) Learning to be thoughtless: Social norms and individual computation. Comput Econ 18:9–24

Falk A, Fehr E, Fischbacher U (2005) Driving forces of informal sanctions. Econometrica 73(6):2017–2030

Fehr E, Camerer C (2007) Social neuroeconomics: The neural circuitry of social preferences. Trends Cogn Sci 11(10):419–427

Festinger L, Schachter S, Black K (1950) Social pressures in informal groups. Stanford University Press, Stanford

Friedrichs J, Blasius J (2003) Social norms in distressed neighborhoods: Testing the Wilson hypothesis. Hous Stud 18(6):807–827

Gibbs J (1966) Sanctions. Soc Probl 14:147–159

Goldberg D (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Boston

Gürerk O, Irlenbusch B, Rockenbach B (2006) The competitive advantage of sanctioning institutions. Science 312:108–111

Hackman J (1992) Group influences on individuals in organizations. In: Dunnette M, Hough L (eds) Handbook of industrial and organizational psychology, vol 3, 2nd edn. Consulting Psychologists Press, Palo Alto, pp 199–267

Hardin G (1968) The tragedy of the commons. Science 162:1243–1248

Hardin R (1995) One for all: The logic of group conflict. Princeton University Press, Princeton

Harrison J, Lin Z, Carroll G, Carley K (2007) Simulation modeling in organizational and management research. Acad Manag Rev 32(4):1229–1245

Hechter M, Opp K-D (eds) (2001) Social norms. Russell Sage Foundation, New York

Heckathorn D (1990) Collective sanctions and compliance norms: A formal theory of group-mediated social control. Am Sociol Rev 55(3):366–384

Henrich J, Boyd R (1998) The evolution of conformist transmission and the emergence of between-group differences. Evol Hum Behav 19:215–241

Henrich J, Boyd R (2001) Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. J Theor Biol 208:79–89

Henrich J, Gil-White F (2001) The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. Evol Hum Behav 22(3):165–196

Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, Bolyanatz A, Cardenas J, Gurven M, Gwako E, Henrich N, Lesorogol C, Marlowe F, Tracer D, Ziker J (2006) Costly punishment across human societies. Science 312:1767–1770

Hill R, Dunbar R (2003) Social network size in humans. Hum Nat 14(1):53–73

Ho A, Leung K (1998) Group size effects on risk perception: A test of several hypotheses. Asian J Soc Psychol 1:133–145

Horne C (2000) Community and the state: The relationship between normative and legal controls. Eur Sociol Rev 16(3):225–243

Horne C (2001a) The enforcement of norms: Group cohesion and meta-norms. Soc Psychol Q 64(3):253–266

Horne C (2001b) Sociological perspectives on the emergence of social norms. In: Hechter M, Opp K-D (eds) Social norms. Russell Sage Foundation, New York, pp 3–34

Horne C (2004) Collective benefits, exchange interests, and norm enforcement. Soc Forces 82(3):1047–1062

Horne C (2007) Explaining norm enforcement. Ration Soc 19(2):139–170

Horne C, Cutlip A (2002) Sanctioning costs and norm enforcement. Ration Soc 14(3):285–307

Hornsey M, Oppes T, Svensson A (2002) It's OK if we say it, but you can't: Responses to intergroup and intragroup criticism. Eur J Soc Psychol 32:293–307

Houser D, Xiao E, McCabe K, Smith V (2008) When punishment fails: Research on sanctions, intentions, and non-cooperation. Games Econ Behav 62(2):509–532

Hurwicz L (2008) But who will guard the guardians? Am Econ Rev 98(3):577–585

Isaac R, Walker J (1988) Group size effects in public goods provision: The voluntary contributions mechanism. Q J Econ 103(1):179–199

Kahan D (1997) Social influence, social meaning, and deterrenc. Virginia Law Rev 83:349–395

Kendal J, Feldman M, Aoki K (2006) Cultural coevolution of norm adoption and enforcement when punishers are rewarded or non-punishers are punished. Theor Popul Biol 70:10–25

Knutson B (2004) Sweet revenge. Science 305:1246–1247

Kollock P (1998) Social dilemmas: The anatomy of cooperation. Ann Rev Sociol 24:183–214

Kurzban R, Houser D (2005) Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. Proc Nat Acad Sci 102(5):1803–1807

Laland K, Odling-Smee J, Feldman M (2000) Niche construction, biological evolution, and cultural change. Behav Brain Sci 23:131–175

Lichbach M (1996) The cooperator's dilemma. University of Michigan Press, Ann Arbor

March J (1996) A preface to understanding how decisions happen in organizations. In: Shapira Z (ed) Organizational decision making. Cambridge University Press, New York, pp 9–32

Marques J, Abrams D, Serodio R (2001) Being better by being right: Subjective group dynamics and derogation of in-group deviants when generic norms are undermined. J Pers Soc Psychol 81(3):436–447

Marwell G, Ames R (1979) Experiments on the provision of public goods. I. Resources, interest, group size, and the free-rider problem. Am J Sociol 84(6):1335–1360

McAdams R (1997–1998) The origin, development, and regulation of norms. Mich Law Rev 96(338):343–433

Milgram S (1974) Obedience to authority. Harper & Row, New York

Nakamaru M, Iwasa Y (2006) The coevolution of altruism and punishment: Role of the selfish punisher. J Theor Biol 240:475–488

Oliver P (1980) Rewards and punishments as selective incentives for collective action: Theoretical investigations. Am J Sociol 85:1356–1375

Ostrom E (1990) Governing the commons: The evolution of institutions for collective action. Cambridge University Press, Cambridge

Posner E (2000) Law and social norms. Harvard University Press, Cambridge

Posner R, Rasmusen E (1999) Creating and enforcing norms, with special reference to sanctions. Int Rev Law Econ 19:369–382

Rapoport A (1988) Experiments with N-person social traps II: Tragedy of the commons. J Confl Manag 32(3):473–488

Rock E, Wachter M (2001) Islands of conscious power: Laws, norms, and the self-governing corporation. Univ Pennsylvania Law Rev 149:1619–1700

Sherif M, Harvey O, White J, Hood W, Sherif C (1954/1988) The Robbers Cave experiment: Intergroup conflict and cooperation. Wesleyan University Press, Middletown. Original, 1954. Reprint edn, 1988

Shinada M, Yamagishi T, Ohmura Y (2004) False friends are worse than bitter enemies: 'altruistic' punishment of in-group members. Evol Hum Behav 25:379–393

Stang D (1976) Group size effects on conformity. J Soc Psychol 98:175–181

Tarzi S (2002) International norms, trade and human rights: A perspective on norm conformity. J Soc Political Econ Stud 27(2):187–202

Tata J, Anthony T (1996) Proportionate group size and rejection of the deviate: A meta-analytic integration. J Soc Behav Pers 11(4):739–753

Tomasello M (1999) The human adaptation for culture. Ann Rev Anthropol 28:509–529

Tooby J, Cosmides L (1992) The psychological foundations of culture. In: Barkow J, Cosmides L, Tooby J (eds) The adapted mind: Evolutionary psychology and the generation of culture. Oxford University Press, Oxford, pp 19–136

Tuomela R (1995) The importance of us: A philosophical study of basic social norms. Stanford University Press, Stanford

Voss T (2001) Game-theoretical perspectives on the emergence of social norms. In: Hechter M, Opp K-D (eds) Social norms. Russell Sage Foundation, New York, pp 105–136

Wendel W (2001) Nonlegal regulation of the legal profession: Social norms in professional communities. Vanderbilt Law Rev 54:1955–1982

Williams K (2007) Ostracism. Ann Rev Psychol 58:425–452

Wilson D, Sober E (1994) Reintroducing group selection to the human behavioral sciences. Behav Brain Sci 17:585–654

Young HP, Foster D (1991) Cooperation in the short and in the long run. Games Econ Behav 3:145–156

Michael J. Prietula is a Fellow in the Institute for Advanced Policy Solutions, Adjunct Professor of Psychology, and Professor in the Goizueta Business School at Emory University. He is also Visiting Scientist at the Institute for Human and Machine Cognition in Pensacola Florida. He has held faculty positions at Dartmouth College, Carnegie Mellon, and Johns Hopkins University. He received his Ph.D. in Information Systems from the University of Minnesota. He is past president of the North American Association for Computational and Social Sciences (NAACSOS).

Daniel Conway is an Associate Professor of Business Administration at Augustana College. He has held faculty positions at the University of Notre Dame, Indiana University, and the University of Florida. His research interests include Information Economics, Game Theory, and Business Process Management. He received his Ph.D. in Decision Sciences from Indiana University. He authors a bi-monthly column for IEEE Security and Privacy magazine.